Constructional tolerance:

Perek, Florent; Hilpert, Martin

DOI:
10.1075/cf.6.2.06per

Document Version
Peer reviewed version

Citation for published version (Harvard):
Perek, F & Hilpert, M 2014, 'Constructional tolerance: Cross-linguistic differences in the acceptability of non-conventional uses of constructions', *Constructions and Frames*, vol. 6, no. 2, pp. 266-304. https://doi.org/10.1075/cf.6.2.06per

Constructional tolerance

Cross-linguistic differences in the acceptability of non-conventional uses of constructions

Florent Perek & Martin Hilpert

Abstract

The present paper investigates the question whether different languages can be categorized into 'constructionally tolerant' languages, which grant speakers considerable freedom to combine syntactic constructions with lexical items in non-conventional ways, and 'valency-driven' languages, which impose stronger restrictions on the way in which constructions and lexical items can be combined. The idea of such a typological distinction is sketched for instance by Rostila (2014). In order to explore possible effects of constructional tolerance, a grammaticality judgment task is administered to speakers of English and French, which are two languages that differ with regard to this phenomenon: English verbs can be used across different argument structure constructions with relative ease, French verbs are more constrained. Both populations of speakers are exposed to stimuli sentences of varying creativity in a second language, namely German. The paper advances the constructional tolerance hypothesis, which states that speakers of a constructionally tolerant language should judge non-conventional examples in an L2 with more lenience than speakers of a valency-driven language. The experimental results are in line with this hypothesis, but they also suggest that grammaticality judgments are influenced by the availability of a productive L1 construction that shows functional overlap.

Keywords: creativity, argument structure constructions, second language, English, German, French, typology

1. Introduction

Construction Grammar (Fillmore et al. 1988; Goldberg 1995, 2006) is increasingly applied to languages other than English (cf. Boas 2010; Ziem & Lasch 2013); further, there exists work that explicitly addresses its potential to operate cross-linguistically and to serve as a flexible framework that is suitable for the empirical description of just about any kind of language (Croft 2001; Fried & Östman 2004). While we fully agree with the claim that Construction Grammar affords a useful descriptive framework regardless of whatever language is studied, we would like to investigate in this paper whether constructional generalizations, especially generalizations in the form of argument structure constructions, are used to similar extents in different languages. More precisely, is it the case that languages can be characterized as giving their lexical items greater or lesser freedom to occur within different morpho-syntactic environments? If so, one might be able to distinguish between languages with regard to their readiness to allow creative combinations of argument structure constructions and lexical elements, specifically lexical verbs. An idea along these lines is in fact brought up by Rostila (2014: 148); some languages may have what we call in this paper

'constructional tolerance', whereas other languages impose a greater degree of morpho-syntactic fixedness on their lexical items. Our initial curiosity regarding this question is rooted in the anecdotal observation that English may actually be typologically unusual in its tolerance for new and unusual combinations of verbs and syntactic frames. In the examples below, verbs occur in syntactic frames that differ from their respective canonical argument structures.

(1)	*Example sentence*	*Construction*
a.	John sneezed the napkin off the table.	caused motion
b.	Mary poured John another whisky.	ditransitive
c.	Emeril sliced and diced his way to TV stardom.	*way*-construction
d.	The truck rumbled down the street.	intransitive motion
e.	Pat kissed Bill unconscious.	resultative
f.	Kate hit at the wasp.	conative

The verb *kiss* (in 1e) encodes a transitive action, but it can be used in the resultative construction to be understood as an action that brings about a resultant state. The fact that speakers can insert a verb such as *kiss* into the resultative construction reflects a global characteristic of English: its verbs and constructions can be combined in novel ways with a relatively high degree of flexibility. It is this tolerance that motivates the recognition of so-called argument structure constructions (Goldberg 1995), which in turn has served as a strong argument for adopting Construction Grammar as a framework for grammatical description more generally. Such phenomena are modelled in various ways in the construction grammar literature, and have been discussed under many different names, including accommodation (Goldberg 1995), coercion (Michaelis 2005, Lauwers and Willems 2011), and type shifting (De Swart 1998, Michaelis 2005), to name only but a few.

Other languages than English, even genetically closely related ones, appear to be less tolerant towards the loose combinability of verbs and constructions. For instance, the following examples show that in a case where the English verb *spray* appears across a range of syntactic frames, the corresponding French source examples involve several different lexical verbs:

(2) a. De la peinture gicla dans l'air.
of the paint sprayed in the air
'Paint sprayed into the air.'

b. Il projeta de la peinture sur le mur.
he sprayed of the paint onto the wall
'He sprayed paint on the wall.'

c. Il aspergea le mur de peinture.
he sprayed the wall of paint
'He sprayed the wall with paint.'

Of course, anecdotal evidence of this kind cannot settle the question whether a language such as French truly is less tolerant than English towards the usage of novel combinations of verbs and constructions. There is always the possibility that the chosen anecdotal examples do not accurately reflect how the language as a whole utilizes argument structure constructions. There could be productive argument structure constructions elsewhere. So how could the constructional tolerance of a language be tested? One strategy would be to investigate the productivity of argument structure constructions in different languages on the basis of comparable corpora. If mutually

corresponding sets of lexical elements are compared with regard to their distributions across different morpho-syntactic frames, it might turn out that one language shows consistently wider distributions of its lexical elements than another. Problems that would have to be addressed in such an approach concern the selection of appropriate lexical items, their matching across languages, and of course the selection of corpora that exhibit similar characteristics in terms of genre, register, and even topic, since lexical elements are at stake. Another analytical strategy would be to test the constructional tolerance of individual speakers in an experimental setting. If different populations of speakers are confronted with a variety of stimuli from their respective language, ranging from fully conventional patterns to creative usages and further to highly marginal examples, differences between languages might emerge. The problem here is that it is virtually impossible to construct stimuli that are comparable in their relative degree of acceptability across entirely different languages.

In order to avoid problems of this sort, our approach in this paper is to use stimuli from a single language, German, in combination with different populations of speakers that acquire German as a second language. In accordance with SLA research on transfer (e.g. Jarvis & Pavlenko 2008), we hypothesize that the L1 of a learner has an effect on how sentences in the target L2 are judged. As will be discussed in more detail below, we compare L2 learners with English and French as their respective native languages in order to test whether English-speaking learners of German exhibit a higher degree of constructional tolerance than French-speaking learners of German. Granting for the moment the idea that there indeed could be systematic cross-linguistic differences in the extent to which grammars make use of argument structure constructions, why would such differences be of interest? The extent to which a language makes use of argument structure constructions is an informative, and as yet unexplored typological parameter that possibly correlates with other structural characteristics. Rostila's (2014) suggestion that languages can be classified as either construction-driven or valency-driven with regard to their use of argument structure opens up a new perspective on cross-linguistic variation. Seeing how such a parameter cross-cuts or aligns with other parameters would not only deepen our understanding of linguistic typology in general, but it would also yield implications for the theory of Construction Grammar and its cross-linguistic application. We thus hold that this distinction is a topic that deserves further scrutiny.

The remainder of this paper is organized as follows. Section 2 presents the experimental methodology that we apply to investigate the question of constructional tolerance. Section 3 discusses the results. Section 4 takes a step back and discusses the general implications of our results for Construction Grammar in the context of stating cross-linguistic generalizations.

2. Methodology

In this section, we describe the experiment that we used to test for potential differences in constructional tolerance between different languages, in our case between English and French. We first present the general method, namely the collection of acceptability judgments, and we motivate our choice to apply this method to sentences in a language that is an L2 for our participants, instead of sentences in their native language. After discussing the stimuli that we used, we describe the experimental procedure and present the groups of participants that were tested.

2.1 Collecting acceptability judgments

The notion of constructional tolerance captures how speakers of a language judge sentences that deviate from conventional usage because they contain a lexical item in an unusual morpho-syntactic context. This leads to the perception of these sentences as creative, deviant, or even ungrammatical. Speakers of a language with high constructional tolerance will more readily 'tolerate' unusual sentences, whereas speakers of a valency-driven language with lower constructional tolerance will be more critical. The degree of constructional tolerance of a given language can thus be measured by collecting acceptability judgments from speakers over sets of sentences instantiating the same construction with different verbs, including sentences that exemplify unusual combinations. There are good arguments for seeing acceptability judgments as gradable rather than binary (Schütze 1996: 62). It is therefore preferable in such a task to let participants choose a value on a scale (either discrete or continuous), whose upper bound indicates full acceptability and whose lower bound indicates full unacceptability, with intermediate degrees of acceptability in between. To illustrate this method, let us consider a German argument structure construction that we call the reflexive-motion construction, as exemplified by (3).

(3) Die Helfer wühlen sich durch den Schutt.
the helpers dig themselves through the rubble
'The helpers are digging their way through the rubble.'

Formally, the construction consists of a nominative subject, a verbal predicate, a reflexive accusative pronoun co-referential with the subject, and a locative complement describing a path. Usually the path is expressed by a prepositional phrase with a complement in the accusative case. The meaning of this construction is very close to that of the English *way*-construction (*She typed her way to a promotion*, cf. Goldberg 1995; Israel 1996), in that it refers to an event in which the subject referent acts in the way described by the verb, and as a result moves along the path referred to by the locative complement. Example (3) provides a typical and perfectly acceptable instantiation of the reflexive-motion construction. If acceptability judgments were to be collected on this sentence, it is expected that the ratings given by native speakers would be on average very high. Like the *way*-construction, the reflexive-motion construction is quite productive in German, yet some examples turn out to be more felicitous than others. For instance, in example (4) below, the construction is used in a clearly more creative way.

(4) Die Ratte nagt sich in die Speisekammer.
the rat gnaws itself in the pantry
'The rat gnaws its way into the pantry.'

While the sentence is not outright ungrammatical, it is expected that acceptability judgments would be worse than in the case of example (3), with some speakers finding it quite acceptable, and others being not so sure about it. Finally, example (5) is arguably highly deviant and would be likely to receive negative judgments from most native speakers.

(5) Der Dieb raubt sich ins Gefängnis.
the thief robs himself into-the prison
'The thief robs his way into prison.'

In sum, an acceptability judgment task seems to be a suitable method to gauge constructional tolerance. Hence, one possibility would be to run an experiment that

confronts speakers of different languages with more or less creative examples in their respective native tongue, and to see whether behavioral differences between these groups emerge. The problem with this approach, as pointed out above, is that such a comparison does not involve a *tertium comparationis*, as it is highly problematic to find linguistic units that form one-to-one matches across two or more languages. It goes without saying that a comparison of the productivity of constructions that fulfill different functions in the target languages (such as the English ditransitive construction, e.g., *John built the kids a merry-go-round*, vs. the French reflexive construction, e.g., *La porte s'est ouverte* 'The door opened') would make little sense with respect to the present investigation. Differences in the meanings of these constructions result in different constraints regarding their productivity, which means that their respective productivity domains are simply not comparable. It could be possible to avoid this problem, or at least attenuate its effects, by using constructions that form translational equivalents. However, a direct translational equivalent in a language A cannot always be found for any given construction in a language B. For instance, to the best of our knowledge, no clause-level construction in French corresponds to the English *way*-construction, so that the construction is typically translated into French in a roundabout way. In other cases, a construction in a language A might map onto more than one construction in another language B (Aijmer & Hasselgård 2004). For example, the English conative construction (V-*at*-NP, e.g., *The lumberjack hacked at the fallen tree*) maps onto two translational equivalents in German, namely the *nach*-conative construction (e.g., *Er schlug nach dem Mann* 'He hit at the man') and the *an*-conative construction (e.g., *Die Mäuse nagten an dem Kabel* 'The mice gnawed at the cord'; cf. Frense & Bennett 1996; Proost 2009). Besides, even when direct and systematic translational equivalents can be identified, the assumption that their productivity domains are similar does not necessarily follow, especially if the target languages are not closely related. Cross-linguistic studies of the constructicon (e.g., Boas 2010, 2011) provide countless other examples of such cases; see in particular Bäckström, Lyngfelt and Sköldberg (2014) for a general comparison between English and Swedish.

To avoid the methodological problems associated with comparisons of acceptability judgments taken from different languages, we propose an alternative approach consisting in testing speakers of different languages with sentences from another language in which both groups of tested speakers are proficient. With this method, the linguistic material presented to participants is held constant across groups, and only the participants' L1 varies. The way speakers use a second language tends to be influenced by their first language, a fact referred to as language transfer in L2 acquisition research (Gass & Selinker 1992). Along the same lines, we hypothesize that the constructional tolerance that speakers acquire with their L1 influences their processing of a second language. In other words, the linguistic feature that we expect speakers to transfer from their L1 to an L2 is the overall knowledge that their language either tolerates or discourages unconventional combinations of lexical items and constructions. Construction Grammar is a theory of linguistic knowledge that models this knowledge as a structured network of symbolic units, in which lexical items and syntactic patterns are represented in the same way (Langacker 1987: 73; Hilpert 2014: 57). Constructional tolerance, in such an understanding of linguistic knowledge, is a global characteristic of that network, which reflects the relative ease with which speakers can create and process new links between lexical items and syntactic patterns. The typological distinction between constructionally tolerant languages and more rigid, valency-driven

languages that we envision is gradient: some languages are very tolerant, some less so, and some not at all.

In the present study, we make the assumption that transfer effects of constructional tolerance should be measurable in an acceptability judgment task. We expect that the acceptability judgments formulated by speakers of different L1s over sentences in a common L2 should be influenced by the relative degree of constructional tolerance found in their respective native language, in that L2 speakers with a more construction-driven L1 should produce more lenient judgments than L2 speakers with a more valency-driven L1. In our experiment, we tested native speakers of English, which is arguably a construction-driven language, and native speakers of French, which appears to be a valency-driven language. Speakers of both groups were exposed to sentences in German, which they were acquiring as one of their second languages. Hence, our study aims to test whether English speakers generally provide more positive acceptability judgments in response to German sentences than French speakers, and especially less negative judgments for sentences that are regarded as deviant by native speakers of German.

2.2 Stimuli

Four German argument structure constructions were selected to create the stimuli sentences: the ditransitive construction, the caused-motion construction, the *nach*-conative construction and the reflexive-motion construction. The reflexive-motion construction was already discussed in the previous section, the three remaining constructions will be briefly described in turn.

The German ditransitive construction: this construction consists of a nominative subject (the agent), a verbal predicate, a dative complement (the recipient) and an accusative complement (the theme), e.g., *[Der Vater]$_{nom}$ zeigt$_{V}$ [seinem Sohn]$_{dat}$ [einen Kartentrick]$_{acc}$* 'The father is showing a card trick to his son'. Like its homologue in many other languages, this construction conveys the idea of a literal or metaphorical transfer and is therefore typically used with verbs of giving (*geben* 'give', *schicken* 'send', *überweisen* 'wire money'), telling (*sagen* 'tell', *beschreiben* 'describe', *erklären* 'explain'), and other related classes.

The German caused-motion construction: this construction consists of a nominative subject (the agent), a verbal predicate, an accusative object (the theme) and a locative complement referring to a source or goal. The latter is usually encoded by a prepositional phrase, with a noun phrase in the accusative for goals, and a noun phrase in the dative for sources. The construction is used to refer to events in which the agent is responsible for a change of location of the theme, e.g., *[Ich]$_{nom}$ hänge$_{V}$ [das Bild]$_{acc}$ [an [die Wand]$_{acc}$]$_{goal}$* 'I hang the picture on the wall'. Typical caused-motion verbs include *stellen* 'put, place', *bringen* 'bring', and *werfen* 'throw'. As in English, many verbs that do not inherently convey caused-motion can nonetheless be used in the construction, in which case the caused-motion meaning is contributed by the construction. Examples such as *[Hans]$_{nom}$ nieste [das Taschentuch]$_{acc}$ [von [dem Tisch]$_{dat}$]$_{source}$* 'Hans sneezed the napkin off the table' are thus fully possible.

The nach-*conative construction*: this construction consists of a nominative subject, a verbal predicate, and a prepositional phrase headed by *nach* (hence with a dative NP complement), e.g., *[Er]$_{nom}$ schlug$_{V}$ nach [dem Mann]$_{dat}$* 'He hit at the man'. It conveys the idea that the action performed by the subject referent is directed towards a second

participant (the referent of the prepositional phrase complement); hence, as with the English conative construction, uses of transitive verbs in the *nach*-construction often suggest that the agent has not reached the second participant and/or has not succeeded in affecting it in a significant way. Typical classes of verbs that can be used in the *nach*-conative construction include verbs of hitting (e.g., *schlagen* 'hit', *treten* 'kick') and verbs of seizing (e.g., *greifen* 'seize, grasp', *schnappen* 'grab').

Table 1 below presents a summary of the constructions used for our stimuli sentences. For each construction, two example sentences are provided with their English translations. One of these sentences exemplifies a fully acceptable use, the other an unacceptable one.

Ditransitive construction: NP_{nom} V NP_{dat} NP_{acc}
Acceptable: Der Vater schenkt ihm eine Bohrmaschine. 'The father gives him a drill (as a present)'
Unacceptable: Der Professor denkt uns eine Erklärung. 'The professor thinks us an answer'
Caused-motion construction: NP_{nom} V NP_{acc} $Prep_{Loc}$ NP_{acc}
Acceptable: Die Retter brachten den Mann in eine Klinik. 'The paramedics brought the man in a hospital'
Unacceptable: Das Kind trinkt den Saft in den Bauch. 'The child is drinking the juice into his belly'
***Nach*-conative construction:** NP_{nom} V *nach* NP_{dat}
Acceptable: Der Spieler tritt nach dem Ball. 'The player kicked at the ball'
Unacceptable: Der Gärtner gießt nach der Pflanze. 'The gardener waters at the plant'
Reflexive-motion construction: NP_{nom} V Refl $Prep_{Loc}$ NP_{acc}
Acceptable: Die Helfer wühlen sich durch den Schutt. 'The helpers are digging their way through the rubble'
Unacceptable: Die Pflanze wächst sich ins Fenster. 'The plant grows its way into the window'

Table 1: German argument structure constructions used for the stimuli sentences.

The second author, a native speaker of German, designed 25 sentences for each construction, aiming to create stimuli that range from typical to creative to downright unacceptable, such that the sentences of each set should receive a wide range of judgments on the acceptability scale. A full list of the 100 stimuli sentences, including the examples given in Table 1, is provided in Appendix A.

2.3 Procedure and participants

The 100 stimuli were presented in randomized fashion to each participant. Each sentence was displayed on a computer screen for 4000 msec, then the sentence disappeared and the participant was asked to rate the sentence in terms of acceptability

with the question “Wie bewertest du den Satz?” (“How do you evaluate this sentence?”). As indicated previously, the acceptability judgments were formulated against a graded scale. We chose a continuous scale, instead of the traditional 7-point Likert scale, which means that participants were able to formulate their judgments by selecting any position on the scale, as opposed to pre-defined discrete graduations. The scale was presented to participants as a bar filled with colors shading horizontally from green on the left, indicating full acceptability, to red on the right, indicating full unacceptability, through shades of yellow and orange, where more mixed judgments were to be placed. This scale is reproduced in Figure 1 below. The arrow and labels are included here for the sake of clarity, but they were not shown in the scale presented to the participants.

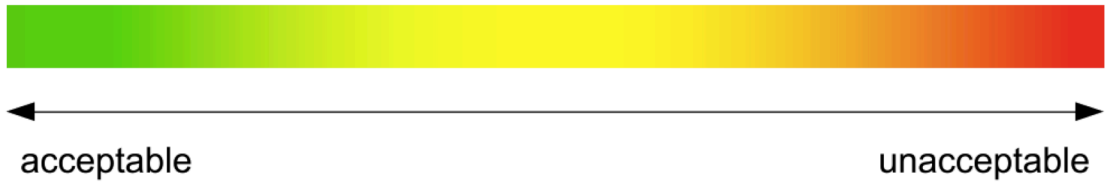

Figure 1: Continuous acceptability scale presented to the participants.

To provide their ratings, the participants had to click with the mouse on the position on the scale where they located each sentence in terms of acceptability. The horizontal coordinate of the mouse cursor at the click was recorded as a measure of the dependent variable in each trial. To familiarize the subjects with the scale and more generally with the task of positioning items on it with respect to some defined criterion, we presented them with a similar assignment prior to the acceptability judgment task, in which they had to rank a range of physical objects (a pencil, a newspaper, a hammer, etc.[i]) on the same green-to-red scale according to their weight, with lighter objects being positioned towards the green side and heavier objects positioned towards the red side. After this warming-up, the sentence acceptability judgment task proper began. The participants received an oral introduction (in German) to the task, which was supported by practical instructions (also in German) on the screen throughout the experiment. The participants were told that they were going to perform a rating task similar to the warming-up sequence, but this time with sentences instead of concrete objects. They were first prompted to form their own idea of what property of sentences they were asked to rate by being shown two sets of sentences: one set was said to be ‘green’ with reference to the scale, the other was said to be ‘red’. These sentences are reproduced in Table 2 below, with their closest translation in English, which was not shown to the participants.

Diese Sätze sind grün: ('These sentences are green')	Diese Sätze sind rot: ('These sentences are red')
Das Kind läuft zur Schaukel. ('The child runs to the swing') Der Angler fängt einen Fisch. ('The fisherman catches a fish') Das Auto bremst. ('The car brakes') Die Schülerin sagt die Antwort. ('The pupil says the answer') Die Touristen kommen nach Freiburg. ('The tourists come to Freiburg')	Das Kind träumt zur Schaukel. ('The child dreams to the swing') Der Angler fängt. ('The fisherman catches') Das Auto bremst zur Straße. ('The car brakes to the street') Die Schülerin sagt dem Lehrer. ('The pupil says to the teacher') Die Touristen besuchen nach Freiburg. ('The tourists visit to Freiburg')

Table 2: Example sentences shown to the participants to illustrate the notion of sentence acceptability.

The subjects were generally able to tell the difference between the two kinds of sentence, in that they felt that there was something wrong with the red sentences. As part of the general instructions, we made the judgment criterion explicit by pointing out that the red sentences had one of three shortcomings. Some element was missing, superfluous, or appeared in an inappropriate context. We insisted that the question was not whether the sentences made sense or not. None of the sentences contained combinations of lexical items that would lead to semantic anomalies. Also, the sentences did not contain any mistakes regarding case marking, verb inflection or agreement. The participants were told that they had to assess to what extent the stimuli sentences were good sentences of German. In order to avoid having judgments cluster towards the extreme values of the scale, the participants were encouraged to use the whole range of values covered by the scale. Three practice items of varying degrees of acceptability were presented to the participants,[ii] after which the experiment continued with the randomized 100 stimuli discussed in the previous section.

Three groups of participants were tested: 35 native speakers of English (20 females, 15 males, aged 19-33, 22.1 on average), 30 native speakers of French (20 females, 10 males, aged 18-27, 21.5 on average), and 44 native speakers of German (34 females, 10 males, aged 18-31, 23.0 on average) as a control group. This latter group was included in the experiment in order to provide a baseline against which the acceptability judgments of the non-native speakers could be evaluated. All subjects were students at the University of Freiburg, Germany, either as regular students or in an exchange program. The participants received €5 or course credit for their participation.

The English- and French-speaking participants were L2 speakers of German at varying levels of proficiency. Since acceptability judgments can obviously vary with language proficiency, this factor must be taken into account in the comparison between groups. In order to assess proficiency in the L2, we used a short vocabulary test taken from the DIALANG system (Alderson 2006). DIALANG is a language diagnosis system developed by several European higher education institutions and meant to evaluate the level of skill in a foreign language, including most of the national languages of the European countries, against the Common European Framework for language learning, by means of a computer program that can be downloaded from the DIALANG website.[iii] Prior to the various tasks of the DIALANG test, a vocabulary questionnaire called 'Vocabulary Size Placement Test' is presented to the testees in order to adapt the test material to their

vocabulary level. This questionnaire consists of a list of 75 words, 50 of which in the German version are German verbs from various semantic classes and registers, and the other 25 are nonce verbs. The latter look like German verbs with regards to their phonotactics and their morphology, but they do not exist in any variety of German. A sample of four German verbs and four nonce verbs taken from the test is provided in (6) and (7) below. The German version of the full test is reproduced in Appendix 2.

(6) German verbs: *schwören*, *zermalmen*, *hineinbekommen*, *chiffrieren*
(7) nonce verbs: *schuttern*, *vertrenken*, *herauspasten*, *aggressieren*

By ticking a box, the testees indicate for each word whether they believe it is an actual German word or a made-up. When the testees have submitted their answers, the program tallies the responses, determining how many true words and how many nonce words have been correctly identified as such. The program computes a vocabulary score between 0 and 1000. The score provides an indication of vocabulary size, which can be interpreted in terms of language skill following the descriptions reproduced in Table 3 below.

Level	Score	Description
1	0-100	This level indicates a person who knows a few words, but lacks any systematic knowledge of the basic vocabulary of the language.
2	100-200	This level indicates a very basic knowledge of the language, probably good enough for tourist purposes or "getting by", but not for managing easily in many situations.
3	200-400	People who score at this level have a limited vocabulary which may be sufficient for ordinary day-to-day purposes, but probably doesn't extend to more specialist knowledge of the language.
4	400-600	People who score at this level typically have a good basic vocabulary, but may have difficulty handling material that is intended for native speakers.
5	600-900	People who score at this level are typically advanced learners, with a very substantial vocabulary. Learners at this level are usually fully functional, and have little difficulty with reading, though they may be less good at listening.
6	900-1000	A very high score, typical of a native speaker, or a person with near-native proficiency.

Table 3: Proficiency levels as evaluated by the DIALANG Vocabulary Size Placement Test, with their corresponding score range and description (source: DIALANG test software).

The English- and French-speaking participants performed the German version of the DIALANG Vocabulary Size Placement Test prior to the experiment proper. While the results of such a test are an imperfect measure of language proficiency as a whole, they provide at least a rough indication. The main advantage of the test is that it is practicable. Furthermore, since the test is intended to measure vocabulary size, it should correlate with variation in acceptability judgments, which require knowledge of L2 words and their syntactic combinatorics.

The distribution of vocabulary scores in each group of non-native speakers is plotted in Figure 2 by means of boxplots. As can be seen from this graph, the English- and the French-speaking participants have very similar median values (marked by black stripes

in the diagram; 321.43 vs. 312.39). Most of the participants in both groups score between level 1 and level 4 (with level 1 slightly more populated in the English-speaking group), which, according to the evaluation scale presented in Table 3, shows that they range from beginners to fairly advanced learners. No participant scores like a native speaker (i.e., above 900), as the maximal value in each group is 800 for the English speakers and 840 for the French speakers. In sum, we can be confident enough that most subjects should have been able to understand most of the vocabulary used in the stimuli sentences, while at the same time their still limited experience with German should leave room for influence from their native language.

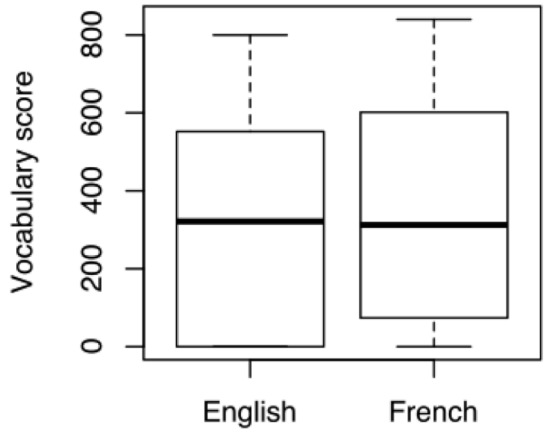

Figure 3: Distribution of vocabulary scores in each group.

In the next section, we turn to an evaluation of the results of this experiment with regards to our hypothesis, i.e., that speakers with different L1s should display systematic differences in their evaluation of sentences in the same L2 in line with the notion of constructional tolerance.

3. Results

In the first section, we present a general analysis of our results by comparing, for each construction, the entire distribution of judgments formulated by each group of speakers. We report that we do not find any systematic effect of constructional tolerance. In the second section, we turn to a more detailed analysis that takes into account the acceptability of sentences as evaluated by native speakers; in this analysis, the dataset is split into 'good' and 'bad' sentences, and within each set of sentences the acceptability judgments formulated by each group of speakers is again compared for each construction. The overall pattern of results that emerge from this analysis seems to be more in line with the constructional tolerance hypothesis, with more noticeable tolerance effects appearing for 'bad' sentences than for 'good' sentences. However, we also find some unexpected differences, which we try to explain and reconcile with the concept of constructional tolerance in the final section.

3.1 General analysis

In this section, we provide a general comparison of ratings provided by each group, for each construction. The distribution of acceptability judgments for each group on sentences of each construction is plotted in the form of a box-and-whisker diagram in Figure 3 below.

This kind of diagram is useful for visualizing how the responses differ across the three languages and the four constructions in the experiment. It can be read as follows. Each

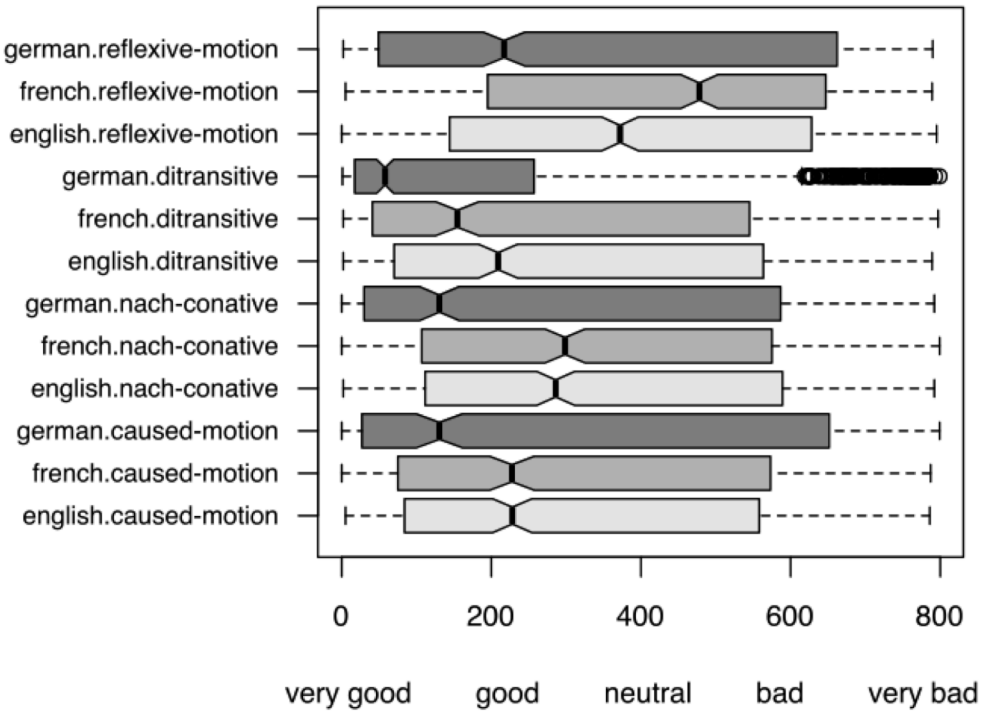

Figure 4: Comparison of acceptability judgments for each construction and each group.

row represents the distribution of acceptability judgments for one group and one construction. The gray box is delimited by the lower and upper quartiles; in other words, it corresponds to the middle range of the distribution and contains half the data. The black stripe is the median value: each half of the distribution is located to the left and right of this value, which can be thus taken as an indication of the central tendency of the distribution. The dashed lines ending with whiskers contain values that are outside the lower and upper quartiles but still within 1.5 times the interquartile range (i.e., the difference between the upper and lower quartiles). These values, while clearly outside of the central tendency of the distribution, are still within a reasonable range of it, but the values that are outside this range are considered as outliers, i.e., values that are abnormally distant from the rest of the distribution. In our dataset, only the distribution of the German ditransitive construction contains outliers; these values are represented in Figure 3 by a line of tightly clustered bullet points towards the right edge of the distribution.

With the exception of the German judgments of the ditransitive sentences, all boxes in the diagram occupy a wide portion of the horizontal axis, which means that the central half of the data points spans a large segment of the grammaticality scale. This is not

surprising given that the stimuli sentences were designed to include uses of each construction with varying degrees of acceptability, ranging from fully acceptable to fully unacceptable. The boxes corresponding to judgments from non-native speakers are slightly shorter (except, again, for the ditransitive construction) and tend to center on higher values (corresponding to lower acceptability) than those corresponding to the native judgments. In other words, the non-native judgments include relatively more middle-range values, as opposed to the extreme ends that mark clear acceptance or rejection. This reveals a higher degree of uncertainty of non-native speakers in judging the acceptability of the German sentences presented to them. For all four constructions, non-native speakers give overall more severe judgments than the native speakers, as the median value of each non-native distribution is higher than the median of the corresponding native distribution.

The main question that our experiment addresses is whether speakers with different L1s (here, English and French) display systematic differences in their judgment of sentences in German. The results presented in Figure 3 do not provide a clear answer to this question, as the difference between the two non-native groups varies substantially according to the construction. For the reflexive-motion construction, English speakers provide more lenient judgments than French speakers. This finding is in line with our hypothesis that English speakers should be more constructionally tolerant than French speakers; however, the same does not hold for the other constructions at all. There is no noticeable difference between groups for the *nach*-conative construction and the caused-motion construction. As for the ditransitive construction, there appears to be a difference, but it is the opposite of that found for the reflexive-motion construction: this time, the English speakers provide harsher judgments than the French speakers.

To test whether the differences between the non-native groups are statistically significant, we submitted our data to a linear regression analysis with mixed effects (cf. Baayen et al. 2008; Baayen 2008: chapter 7). Linear regression is a statistical analysis technique geared towards evaluating the influence of multiple variables (the predictors) on a continuous response (here, the acceptability score for each sentence). In addition to the fixed effects, which correspond to variables manipulated during the experiment, a mixed-effects regression model includes random factors that reflect unsystematic variation that could not be systematically controlled for in the experiment. Typically, random effects capture the variation related to individual subjects (i.e., some subjects display a different behavior than others, regardless of the stimulus) or items (i.e., some items may receive a different response than others regardless of the controlled variables that are manipulated within them). In our model, the fixed effects are Construction (caused-motion, ditransitive, *nach*-conative, reflexive-motion) and Group (English vs. French), as well as the interaction between these terms, to control whether the effects of these predictors may be non-additive (i.e., whether the effect of Group might vary according to the value of Construction). The random effects are Subject and Item. The formula for this model is given in (8) below.

(8) Score ~ Group × Construction + (1|Subject) + (1|Item)

We used the lmer function from the lme4 package (Bates et al. 2011) in the R environment (R Development Core Team 2012). For this regression analysis as well as all subsequent ones, we removed the outliers in the distribution of ratings of each sentence by each group, deleting those values that are greater or lower than 1.5 times the interquartile range. We did this in order to achieve greater statistical power, and while *a priori* data trimming can be questionable in principle, we believe that it is

justified in our case, since we aim to evaluate how different populations of speakers rate particular sentences. In all cases but one (namely, the analysis of ratings given by non-native speakers to the set of 'bad' sentences, discussed in Section 3.2), removing the outliers did not strongly impact the statistical significance of the various predictors, yet reinforced the measured differences between groups. The *p*-values for the fixed effects and their confidence intervals were estimated by means of Markov Chain Monte Carlo sampling with a sample size of 10,000, using the pvals.fnc function from the languageR package (Baayen 2008). The results of the regression analysis are reported in Table 4.[iv]

Predictor	**Estimate**	**MCMCmean**	**Lower**	**Upper**	**pMCMC**	**Pr(>|t|)**
(Intercept)	381.3	381.52	330.86	431.06	0.0001	0
Group: French	**48.07**	**47.77**	**11.6**	**83.71**	**0.0098**	**0.0115**
Construction: ditransitive	**-86.17**	**-85.89**	**-152.44**	**-19.12**	**0.0112**	**0.0312**
Construction: *nach*-conative	-48.06	-47.58	-115.14	14.98	0.154	0.2292
Construction: caused-motion	-78.37	-77.87	-144.33	-11.72	0.0194	0.0552
Group: French × Construction: ditransitive	**-84.53**	**-84.42**	**-112.84**	**-56.71**	**0.0001**	**0**
Group: French × Construction: *nach*-conative	**-52.9**	**-52.71**	**-79.24**	**-23.63**	**0.0002**	**0.0002**
Group: French × Construction: caused-motion	**-48.56**	**-48.47**	**-77.27**	**-20.68**	**0.0008**	**0**

Table 4: Effects of Group × Construction in the linear regression analysis (non-native speakers only). Reference levels: English for Group and reflexive-motion for Construction.

The estimates (in the second column) capture the strength of the effect of each predictor on the model's response; in our case, they correspond to the variation on the acceptability scale that each level of the predictors impacts on the acceptability score, as compared to the intercept, i.e., the value of the acceptability score when all predictors are at their reference levels. A positive value of the estimate indicates that the corresponding level of the predictor produces more severe judgments of acceptability; conversely, a negative value indicates that this level produces more lenient judgments. The *p*-values indicate whether the effect of each predictor (whatever its size) is statistically significant. Statistically significant predictors are indicated by boldface in Table 4.

As can be seen from Table 4, there is a significant main effect of Group, with the French on average giving harsher judgments. There is also a statistically significant effect of the ditransitive construction, and a marginally significant effect of the caused-motion construction, indicating that both groups judge these constructions with more leniency than the reflexive-motion construction, which represents the reference level, i.e. the standard of comparison. However, the same two predictors are also involved in significant interactions, as can be seen in the last three rows of Table 4; in other words, the differences between groups vary significantly according to the construction. To get a clearer idea of how the predicted scores of the two groups vary for each construction, we can sum the estimates of the relevant predictors (among the statistically significant ones) for each Group × Construction combination, thus obtaining the construction-specific differences in acceptability between each group presented in Table 5. A negative difference means that the English are more tolerant towards the sentences of that construction; a positive difference means that the French are.

Construction	Predicted (English)	Predicted (French)	Difference (English - French)
caused-motion	381.3	380.81	-0.49
ditransitive	295.13	258.67	36.46
nach-conative	381.3	376.47	4.83
reflexive-motion	381.3	429.37	-48.07

Table 5: Predicted acceptability scores for each group and each construction.

These results largely confirm our initial observations based on the boxplot diagram. Generally speaking, we do not observe tolerance effects across the board. The difference in acceptability between judgments formulated by each group varies greatly according to the construction. There is virtually no difference with the caused-motion and *nach*-conative constructions, and our hypothesis that English speakers provide more lenient judgments is only borne out for one construction, the reflexive-motion construction. Moreover, for the ditransitive construction, the observed difference is in the opposite direction from that predicted by our hypothesis, since the French are more tolerant with this construction. In conclusion, it seems that this first general analysis does not lend much credence to the constructional tolerance hypothesis.

There are, however, two interpretations of how the hypothesized phenomenon of constructional tolerance is expected to bear on speakers' performance at evaluating sentences in a foreign language. The first, broadest interpretation is that effects of constructional tolerance would arise for any sentence. It is this interpretation that we tested in this section; as we could see, it does not hold. The second, more restricted interpretation is that constructional tolerance effects would arise more noticeably, or perhaps exclusively, for uses of a construction with a lexical item with which it is not conventionally associated. Constructional tolerance could thus show up specifically in reactions towards deviant examples. This more restricted version of the hypothesis seems more reasonable, since regular patterns of the language, including conventional associations between lexemes and constructions, are to some extent likely to be familiar to L2 learners; there is therefore no *a priori* reason why non-native speakers with different L1s should give substantially different judgments on sentences containing such patterns. It might thus be the case that the results reported in this section are due to the 'good' sentences, where constructional tolerance might play a lesser role, and that any effects of constructional tolerance are overshadowed by effects of convergence in acceptability judgments by non-native speakers on clearly acceptable sentences. It remains to be seen whether differences between groups can be observed for those sentences that are judged as deviant by native speakers. We address this question in the next section.

To conclude this section, a few comments concerning the influence of language proficiency are in order. As we explained in Section 2.3, the L2 proficiency of our non-native participants in German was measured by their score on the DIALANG Vocabulary Size Placement Test. In order to test whether speakers of different proficiency levels behave differently when judging the acceptability of German sentences, we included the vocabulary level (VocLevel), as described in Table 3, as an additional predictor in the mixed-effects linear regression model used previously. We conducted separate analyses for the English and French subjects to avoid having a three-way interaction in the regression model (i.e., Group × Construction × VocLevel). The formula for the new model is given in (9) below.

(9) Score ~ Construction × VocLevel + (1|Subject) + (1|Item)

For reasons of space, we do not include the full details of this analysis in this paper. The main findings concerning the effect of proficiency are the following. As with the predictor Group in the previous analysis, the effect of VocLevel consists of significant interactions with the predictor Construction rather than of a single main effect. More proficient speakers in both groups give substantially harsher judgments for the reflexive-motion sentences (i.e., their ratings increase by 26.08 for the English speakers and by 21.31 for the French speakers for each point on the DIALANG proficiency level scale). More proficient English speakers also give harsher judgments for the caused-motion sentences (9.27 increase for each point of the proficiency scale), and more lenient judgments for the ditransitive sentences (11.59 decrease for each point of the proficiency scale). The effects of VocLevel with the other constructions are all negligible (though significant). In sum, while the general trend is that more proficient speakers appear to be more critical of the German sentences, this tendency is only found for some constructions, and the opposite is actually observed for English speakers on the ditransitive construction. Hence, the influence of proficiency cannot be interpreted as a varying effect of constructional tolerance systematically related to the participants' experience with their L2; rather, it probably stems from language transfer effects related to specific constructions in the L1. Because proficiency does not seem to have a consistent effect, and since we are not directly concerned with this factor in this paper, we will no longer consider proficiency in the rest of the analysis.

3.2 Factoring in the native acceptability ratings

In this section, we provide an analysis of our results that takes into account how native speakers judged the acceptability of the sentences in our dataset. As indicated previously, the stimuli sentences were also presented to a control group of native speakers of German. The distribution of their acceptability judgments for each sentence over the acceptability scale is plotted in the leftmost diagram in Figure 4 in the form of boxplots (cf. Section 3.1). Each line corresponds to the distribution of the acceptability ratings for one sentence. The sentences are ordered by increasing median value (corresponding to judgments ranging from more acceptable to less acceptable), from bottom to top. Hence, the vertical dimension can be used to discriminate between 'good' sentences (i.e., judged more acceptable) towards the bottom, and 'bad' sentences (i.e., judged less acceptable) towards the top. The notation is the same as in Figure 3; notably, the boxes and the dashed lines reflect how the distributions of judgments span across the acceptability scale, and thus give an indication of the degree of agreement between speakers for each sentence (which can be seen to vary), with wider boxes (and, to a lesser extent, longer lines) indicating more variability and smaller boxes indicating more consistency. To provide a comparison between groups, two similar diagrams corresponding to the two groups of non-native speakers are also provided.

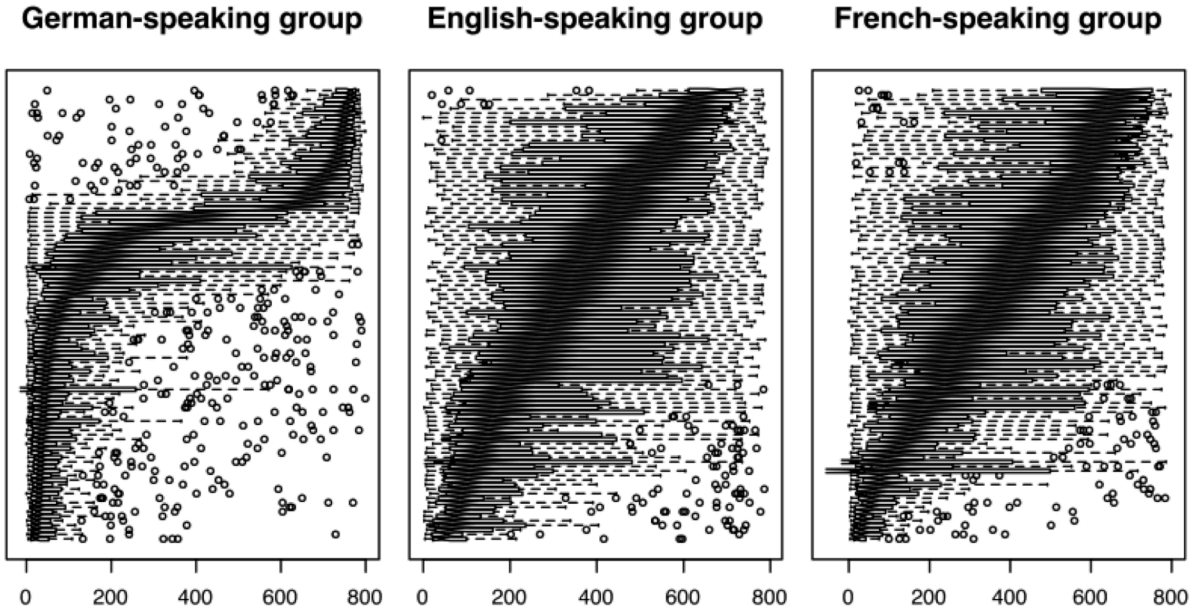

Figure 5: Distributions of acceptability judgments for each sentence and in each group, ordered by their median value.

When comparing the judgments from the native speakers to the judgments of the two learner populations, a striking difference emerges: the German data follow an S-curve, with two clusters of sentences towards each end of the scale, and a relatively narrow area of variable judgments in the middle. The curves of the learner populations form a diagonal, rather than an S-curve. We can also observe that in the judgments of native speakers, the sentences with median acceptability ratings at the extreme points of the scale generally receive fairly consistent judgments across speakers. This is shown by the shorter boxes towards the bottom and the top of the graph. Conversely, judgments of sentences in the middle zone display more variation. This shows that the native speakers are able to identify most sentences in our stimuli set as either clearly acceptable or clearly unacceptable with a high degree of consistency, but they are less certain about the few remaining intermediate cases. Moving on to the learner data, the ordered mean values of the English and French ratings follow a straight diagonal, indicating that the entire scale is equally populated, with judgments being spread out evenly across the available scale. This lines up with our previous observation from Figure 4 that our non-native subjects, irrespective of their native language, appear to be more uncertain about what can or cannot be said in their L2, which results in substantial variation in their ratings for all but a few sentences.

The ratings formulated by the native speakers provide robust data about the acceptability of sentences in our dataset, in which two main groups emerge: on the one hand, we find 65 'good' sentences, with a median acceptability score below 200, and on the other hand, 27 'bad' sentences, with a median acceptability score over 600. The eight remaining sentences, with a score between 200 and 600, will be ignored in the subsequent analysis, due to the lower number of datapoints and, more importantly, the higher variability in the judgments produced by the native speakers. On the basis of this classification, an analysis similar to that presented in the previous section can be performed on each set of sentences independently: the 'good' sentences on the one hand, and the 'bad' sentences on the other. The distribution of acceptability judgments provided by each group for each construction is plotted separately for 'good' vs. 'bad' sentences in Figure 5.

For the 'good' sentences, the comparison of the non-native groups yields results that are similar to ones we found for the whole dataset. English speakers give more lenient ratings to the reflexive-motion sentences, and French speakers give more lenient ratings to the ditransitive sentences. There is no noticeable difference between groups for *nach*-conative sentences. The only difference from our previous findings is that the English

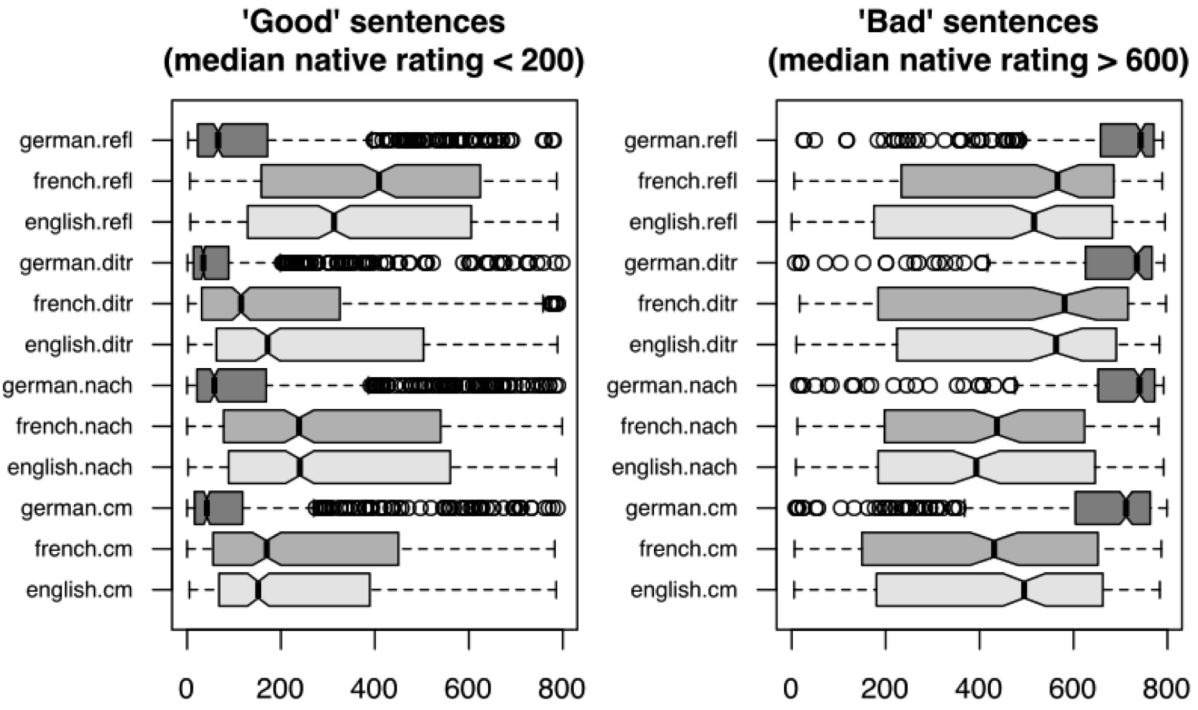

Figure 6: Comparison of the acceptability judgments of 'good' vs. 'bad' sentences for each construction and group.

speakers are slightly more tolerant towards the caused-motion sentences. For the 'bad' sentences, however, a markedly different pattern of results can be observed. The English speakers are still more tolerant towards the reflexive-motion sentences, but the difference in acceptability is weaker. The higher tolerance of French speakers towards ditransitive sentences is no longer found for the 'bad' sentences; instead, the English speakers are slightly more tolerant this time. For 'bad' *nach*-conative sentences, there is a noticeable difference in acceptability between groups, whereas there was none with

'good' sentences. Finally, for caused-motion sentences, the French speakers give more lenient judgments than the English speakers.

As previously, we submitted our data to a linear regression analysis with mixed effects, with the same formula as in (8), to test whether the differences between the non-native groups are statistically significant. The results of the analysis of the 'good' sentences set are presented in Table 6. A systematic comparison of the predicted acceptability scores for each Group × Construction combination is provided in Table 7.

Predictor	**Estimate**	**MCMCmean**	**Lower**	**Upper**	**pMCMC**	**Pr(>|t|)**
(Intercept)	251.56	251.8	200.67	303.57	0.0001	0
Group: French	**-51.08**	**-51.12**	**-85.74**	**-14.45**	**0.0042**	**0.0051**
Construction: *nach*-conative	47.4	46.94	-21.91	118.1	0.2002	0.2625
Construction: caused-motion	-32.48	-32.25	-107.88	42.98	0.4028	0.4674
Construction: reflexive-motion	**98.84**	**98.54**	**24.9**	**177**	**0.0096**	**0.0269**
Group: French × Construction: *nach*-conative	**49.75**	**49.72**	**18.55**	**82.93**	**0.0034**	**0.0025**
Group: French × Construction: caused-motion	**63.62**	**63.67**	**29.25**	**97.72**	**0.0004**	**0.0003**
Group: French × Construction: reflexive-motion	**96.67**	**96.62**	**62.21**	**130.5**	**0.0001**	**0**

Table 6: Effects of Group × Construction in the linear regression analysis of the 'good' sentences set (non-native speakers only). Reference levels: English for Group and ditransitive for Construction.

Construction	**Predicted (English)**	**Predicted (French)**	**Difference (English - French)**
caused-motion	251.56	264.1	-12.54
ditransitive	251.56	200.48	51.08
nach-conative	251.56	250.23	1.33
reflexive-motion	350.4	395.99	-45.59

Table 7: Predicted acceptability scores for each group and each construction in the 'good' sentences set.

As in the boxplot diagram, the results of the regression analysis are similar to those reported for the whole set of sentences. There is again a statistically significant main effect of Group, and there are significant interactions of Group with all levels of the predictor Construction. The comparison of acceptability scores predicted by the model between groups for each construction reveals a significant higher tolerance of the English speakers towards the reflexive-motion construction, and of the French speakers towards the ditransitive construction. The difference found for the *nach*-conative construction, though significant, is negligible. The difference for the caused-motion sentences, already observed in the boxplot, is significant but is markedly smaller than for the first two constructions (12.54 points on the scale vs. 51.08 and 45.59).

The same regression analysis was performed on the set of 'bad' sentences. The results of this analysis are presented in Table 8.

Predictor	**Estimate**	**MCMCmean**	**Lower**	**Upper**	**pMCMC**	**Pr(>|t|)**
(Intercept)	435.771	435.471	352.121	517.43	0.0001	0
Group: French	-31.628	-31.539	-84.908	27.61	0.2672	0.3064
Construction: reflexive-motion	3.694	3.903	-104.47	114.57	0.9336	0.9529
Construction: ditransitive	35.184	35.866	-82.93	164.61	0.5648	0.6213
Construction: *nach*-conative	-23.404	-23.257	-143.149	90.92	0.6868	0.7286
Group: French × Construction: reflexive-motion	**67.476**	**67.425**	**16.949**	**117.83**	**0.0072**	**0.0081**
Group: French × Construction: ditransitive	**52.971**	**53.166**	**-4.164**	**110.54**	**0.065**	**0.0706**
Group: French × Construction: *nach*-conative	28.756	28.783	-24.638	81.93	0.2908	0.2932

Table 8: Effects of Group × Construction in the linear regression analysis of the 'bad' sentences set (non-native speakers only). Reference levels: English for Group and caused-motion for Construction.

In this model, only one predictor turns out to be statistically significant, namely the interaction between Group and Construction for reflexive-motion sentences. For ditransitive sentences, the interaction is only marginally significant. Also, the interaction French × caused-motion is significant when the reference level for Construction is changed to reflexive-motion (p = 0.0081, Estimate = -67.476, MCMCmean = -67.249, Lower = -115.31, Upper = -16.96). No other significant main effects or interactions are found when other reference levels are chosen. A calculation of the predicted scores for each group and construction as in Table 5 and 7 is not useful for these results, since the effects of the predictors amount to interaction effects.

Overall, the model for 'bad' sentences has a weaker predictive power than the model for 'good' sentences, which can largely be attributed to the fact that the 'bad' set contains markedly fewer sentences (27 vs. 65) and hence fewer datapoints. Nevertheless, we are able to measure a significant tolerance of English speakers towards reflexive-motion sentences (of 67.476 points on the scale), a marginally significant tolerance of English speakers towards ditransitive sentences (of 52.971 points), and a significant tolerance of French speakers towards caused-motion sentences (of 67.249 points) when the reference of Construction is manipulated. The visible difference for *nach*-conative sentences observed in the boxplot does not turn out as significant in the regression model, but because of the sparsity of datapoints, it is not possible to conclude that the hypothesis of a behavioral difference between the two groups must be rejected.

To summarize, the general finding of this section is that differences in acceptability judgments between the two non-native groups do vary according to the acceptability of sentences as judged by the native speakers. The English speakers are more tolerant towards both 'good' and 'bad' sentences than the French speakers. The French speakers are relatively more tolerant towards 'good' ditransitive sentences than the English speakers, but English speakers are more tolerant towards 'bad' ditransitive sentences. The English speakers are slightly more tolerant towards the 'good' caused-motion sentences than the French speakers, but they are actually less tolerant than the French towards 'bad' caused-motion sentences. Finally, no substantial difference was found between groups in their judgments of the 'good' *nach*-conative sentences; for the 'bad' sentences, a higher tolerance of English speakers could be observed but was not found statistically significant.

4. Summary and discussion

In the previous section, we presented a quantitative analysis of the results of our experiment. We started with an analysis of the acceptability judgments collected for the whole set of sentences, specifically comparing the distribution of judgments of the two groups of non-native speakers for each construction. The results of this first, general analysis were inconclusive and did not give much credence to the constructional tolerance hypothesis. Only for the reflexive-motion sentences did our subjects behave as expected, i.e. the English speakers were more tolerant towards this construction than the French speakers. For two constructions, the caused-motion construction and the *nach*-conative construction, there was no substantial difference in acceptability judgments between the two groups. For the ditransitive sentences, we did find a difference between groups, which was, however, in the opposite direction from what was expected, since the French speakers were more tolerant towards this construction than the English speakers.

On the basis of these results, we had to reject the more general version of the constructional tolerance hypothesis, whereby constructional tolerance effects are expected to arise for any sentence, resulting in more lenient judgments from constructionally tolerant speakers for fully conventional sentences and more deviant ones alike. We turned to an evaluation of a more specific version of the hypothesis, whereby constructional tolerance effects are expected to arise more particularly with sentences that appear deviant, but not necessarily for sentences that follow conventional construction-lexeme combinations.

To operationalize this more specific version of the hypothesis, we isolated two sets of sentences in the dataset according to the acceptability judgments provided by our group of native speakers, as measured by the median value: on the one hand, 65 'good' sentences falling into the top range of the acceptability scale, on the other hand, 27 'bad' sentences falling into the bottom range. The eight remaining sentences falling into the medium range of acceptability were removed from the dataset. We performed the same analysis as above on the sets of 'good' vs. 'bad' sentences separately, and compared the results.

The main finding in this second analysis is that a different pattern of results does emerge for 'good' vs. 'bad' sentences. The differences between groups for 'good' sentences are very similar to those found in the whole dataset, with significant differences for all constructions. With 'bad' sentences, however, the differences between groups vary substantially. The general trend is that English speakers give more lenient judgments than the French speakers; this trend is found for three of the four constructions. The difference in acceptability is only marginally significant for the ditransitive construction, and not significant for the *nach*-conative construction. The poorer results of the regression analysis might be explained by the relatively low number of 'bad' sentences (as noted, only a subset of our sentences were rejected as 'bad' by our control group). An analysis based on these sentences has to make do with a lower number of observations and therefore suffers from reduced statistical power. At any rate, the results are overall more in line with the predictions of the constructional tolerance hypothesis.

Yet, we also obtain unexpected results that deserve closer examination. First, the higher tolerance of English speakers for reflexive-motion sentences is found both for 'good' and for 'bad' sentences, while our expectation was that there should not be differences between groups for 'good' sentences. Second, the French speakers are more tolerant

towards 'good' ditransitive sentences than the English speakers. This seems to conflict with the expectation that English speakers are more constructionally tolerant, and this is all the more surprising since English speakers actually turn out to be more tolerant towards 'bad' sentences. Third, the French speakers are more tolerant towards 'bad' caused-motion sentences than the English speakers, which again runs counter to expectation, as the French speakers are expected to be less constructionally tolerant in general, and in particular since the French caused-motion construction is arguably less productive than the English caused-motion construction. Moreover, the opposite tendency was found for 'good' sentences.

As it turns out, these unexpected findings can be explained by characteristics of the relevant construction with respect to its counterpart in the L1 of our subjects, or can be argued to be compatible with the constructional tolerance hypothesis after all. In the remainder of this section, we address each of these points in turn.

4.1 Why are the English speakers tolerant towards the 'good' reflexive-motion sentences as well as the 'bad' ones?

This question receives a sensible answer when one looks at the reflexive-motion construction more closely and compares it to the other constructions in our dataset. The other constructions are taken from the inventory of basic argument structure constructions in German, especially the ditransitive and the caused-motion construction, and therefore represent regular ways of verbal argument encoding. That is not the case with the reflexive-motion construction, which, not unlike the English *way*-construction, represents more creative, figurative language. This difference in discursive function may explain why the reflexive-motion construction is less sensitive to conventionality than the other constructions, and thus more open in terms of the verbs that can be used in it. Consequently, only strikingly deviant uses of the construction are rejected by native speakers, for instance those with verbs of self-propelled motion (e.g., *laufen* 'walk, run', *schweben* 'float', *steigen* 'climb, get on', *tauchen* 'dive'). Speakers may judge these verbs as semantically incongruent with the construction, since the construction is a tool for recruiting non-motion verbs into the service of expressing 'motion by means of a non-motion activity'. The unacceptability of such cases is so obvious that both groups of non-native speakers are on a par in their judgment, which reduces the effect of constructional tolerance. However, sentences considered acceptable by native speakers are less clear to the non-natives: this is where constructional tolerance comes in, with the less constructionally tolerant French being thrown off by such sentences and the English being more prone to accept them. The effect is also probably reinforced by the fact that similar sentences with a non-subcategorized reflexive pronoun are possible in English (e.g., *He drank himself under the table*), but not in French (**Il s'est bu sous la table*).

4.2 Why are the French speakers more tolerant towards the 'good' ditransitive sentences than the English speakers?

This question can be answered with regard to the kinds of ditransitives that are found in each set, and how they compare to ditransitives in English and French. In general, the set of 'good' German ditransitives includes uses that have direct counterparts in French but not in English, for example malefactives (*stehlen* 'steal', *rauben* 'rob'), benefactives that would be better translated in English with a *for*-PP (*Der Klemptner schickt uns seinen Kollegen* vs. *?The plumber sends us his colleague*, cf. Bresnan et al. 2007 for evidence that

animate themes are strongly dispreferred in the ditransitive construction), and ditransitive uses of various verbs whose English counterparts must be used with a prepositional *to*-dative (e.g., *erklären* 'explain', *spenden* 'donate', *widmen* 'dedicate'; cf. *He explained the problem to him* vs. **He explained him the problem*). Another factor making English speakers shy away from the 'good' ditransitive sentences might be the presence of so-called inalienable possession ditransitives, where the recipient argument is understood as the possessor, broadly construed, of the theme argument, as in *Das Gericht erlaubte ihr die Abreise* 'The court allowed her departure' (lit. 'The court allowed her the departure'). Both German and, to a lesser extent, French possess this kind of structure, which English simply does not have. In sum, the French and German ditransitive constructions appear to show substantial areas of functional overlap, whereas the usage of the English ditransitive construction is different in many respects. It thus comes as no surprise that the English speakers tend to reject many of the correct German ditransitives, whereas the French speakers see no problem with them. Conversely, the ditransitive sentences rejected by the native speakers exemplify uses whose counterparts in either English of French are equally deviant, e.g., the atypical use of a verb of sound emission to describe the transfer of a message in *Der Moderator lacht ihm die Antwort* '?The TV presenter laughs him the answer'. Yet, a difference can be observed between the two non-native groups, which lines up with the assumed higher constructional tolerance of English. In sum, while similar idiosyncrasies of the French and German ditransitive construction appear to cancel out any role that constructional tolerance might have, the effects of constructional tolerance rise again in those cases where such correspondences are not available.

This explanation also accounts for the finding that more proficient English speakers (but not more proficient French speakers) are more tolerant towards the ditransitive sentences than less proficient speakers, as reported in Section 3.1: as their competence in German increases, language transfer effects from English decrease, and the English speakers learn to accept uses of the ditransitive construction in German that do not have a direct equivalent with the English ditransitive construction.

4.3 Why do the French speakers appear to be more tolerant towards 'bad' caused-motion sentences than the English speakers?

One tentative explanation for this question relates to the fact that French does not have a caused-motion construction in the same sense that English and German have. This is largely, though probably not only, due to the fact that French possesses more limited ways to unambiguously encode motion in the prepositional phrase. Most locative prepositions in French are ambiguous between a location and a path interpretation, contrary to English, where for example *into* and *onto* invariably convey motion. As a result, the locative PP in many sentences whose literal translation in English or German could receive a caused-motion interpretation is strongly biased towards a location interpretation in French. For instance, *Elle promène le chien sur la place*, literally 'She walks the dog on/onto the square', would usually be understood as describing an event of walking a dog within a certain area (the square), but not in a caused-motion interpretation profiling the motion of the dog along the whole path leading into this area. This contrasts with the example *Elle promène le chien vers la place*, in which the PP with *vers* 'towards' unambiguously describes a path.

If French speakers are more familiar with location PPs than path PPs following a direct object, this could bias them towards interpreting such PPs as locations in German

caused-motion sentences, especially with such ambiguous prepositions as *in* 'in' or *auf* 'on' and with verbs that are not normally used in caused-motion sentences. Through this misinterpretation, 'bad' sentences would receive better acceptability. Conversely, greater familiarity with this kind of construction would lead speakers of English towards an interpretation of the PP as a path, which would yield the correct intended sentence meaning, and, ultimately, less positive judgments. Admittedly, the accusative case marking in our test sentences prevents a location interpretation in principle, but in the context of our experiment, it is possible that, under the time pressure, our French participants did not always notice the accusative marker or did not automatically infer that the locative PP referred to a path of motion, or ignored case marking altogether.

This explanation seems particularly satisfactory in the case of two sentences regarding which the tolerance of French speakers was markedly higher than that of English speakers: *Der Müllmann greift den Sack in den Müllwagen* 'The trash collector grabs the bag and drops it in the garbage truck', and *Der Einbrecher droht das Opfer ins Badezimmer* 'The burglar threatens the victim into the bathroom', in which the locative PPs could have been interpreted as modifiers of the object NP rather than arguments of the verb, i.e., 'the bag that is in the garbage truck' and 'the victim that is in the bathroom'. For two other sentences, that explanation seems less likely: *Der Hund bellt die Kinder ins Haus* 'The dog barks the kids into the house' and *Das Kind trinkt den Saft in den Bauch* 'The child drinks the juice into his belly'. In the former, the combination *bellen* + NP ('bark so./sth.') is not a well-formed predication, regardless of whether the following PP is taken as a NP modifier or a sentence adjunct. In the latter, *Bauch* 'belly' is an unlikely location for the whole event to take place, or for the juice to originally come from before being drunk. Yet, there are two reasons why these two cases are not necessarily at odds with our explanation. First, the judgments of each group for these are less far apart than for the first two sentences. Second, we observe that this difference decreases with the vocabulary level (even changing directions for level 5 speakers and the *bellen*-sentence), which seems to indicate that the observed difference in acceptability might have to do with misunderstanding due to vocabulary gaps. At all events, our explanation is also compatible with the finding that the 'good' caused-motion sentences receive more lenient judgments from the English speakers than from the French speakers. Most of the 'good' caused-motion sentences contain a verb that typically describes caused-motion, such as *bringen* 'bring', *stecken* 'put, tuck' or *werfen* 'throw', or easily accept a caused-motion interpretation, such as *tragen* 'carry' or *ziehen* 'pull'. Sentences with such verbs are therefore readily interpreted as instances of the caused-motion construction by both groups of non-native speakers, which should not incur substantial differences in acceptability. Since there are fewer sentences with a verb that does not normally describe caused-motion (the only three relevant verbs – out of 14 – being *bestellen* 'order, summon', *bitten* 'beg, request' and *pfeifen* 'whistle'), which could confuse the French speakers into a location interpretation of the PP, the result that English speakers show an overall slightly higher tolerance for the 'good' caused-motion sentences is not unexpected.

5. Conclusion

Summing up, the results obtained in our study yield evidence that is in line with the constructional tolerance hypothesis in its more restricted form, namely as tolerance towards recognizably creative combinations of lexical items and argument-structure constructions. We hasten to add that the effects of constructional tolerance are filtered

by the constructions that exist in learners' L1, which are used as reference points in the construction of what could be called 'diasystematic' generalizations (cf. Höder 2012). This means that while speakers of English exhibit constructional tolerance as a general tendency, their tolerance towards uses of a specific L2 construction varies with the availability of a productive L1 construction that shows functional overlap. This finding suggests that constructional tolerance might not fundamentally be a property of a language *per se*, but rather, first and foremost, one that follows from the constructions found in that language. A language can be described as constructionally tolerant to the extent that it has many constructions that allow creative uses. English can be considered typologically peculiar in this respect for appearing to have a particularly high number of such constructions.

Note, however, that the apparent construction-specificity of tolerance effects does not in principle preclude the possibility that speakers truly exhibit some form of constructional tolerance across the board, emerging from the knowledge that their language tends to allow creative combinations (or conversely). This idea is actually in line with experimental findings of artificial language learning studies. For example, Wonnacott et al. (2008) had subjects learn an artificial language containing nonce verbs used in two different but interchangeable novel constructions. In the input of scene-sentence pairings given to participants, verbs either occurred in only one construction, or alternated between both constructions. Wonnacott et al. find that participants were more likely to use a verb productively (viz. in a construction with which it was not attested) if the language contained many alternating verbs, which leads them to conclude that "learners are sensitive to statistical information above the level of individual verbs" (p. 204). Wonnacott (2011) replicate these results with children, and Perek and Goldberg (submitted) report similar findings with constructions that differ in pragmatic function. In other words, participants in Wonnacott et al.'s (2008) experiment not only learned the behavior of individual verbs, they also appeared to be aware of the extent to which the syntactic distribution of verbs tend to be constrained in general, which is one aspect of the notion of constructional tolerance put forward in this study.

Importantly, if effects of constructional tolerance are related to the specific constructions of a language, this also means that they are likely to be highly domain-specific; in other words, a language may appear constructionally tolerant when realizing certain functions, but not others. Correspondingly, language learners might exhibit levels of constructional tolerance in their L2 in a similar context-dependent way; this is precisely what we found in our experiment. In order to understand the effects of constructional tolerance more fully, it would therefore be desirable to design studies with a wider range of constructions, to counter the fact that the reported effects vary according to the construction. In particular, future studies should focus on constructions which lack a clear translation equivalent in the L1s of participants (or a counterpart with parallel syntactic properties), though ones that participants are likely to be familiar with or find relatively transparent. Possible candidates (for English and French participants) could include the *mit*-predicative construction (Hilpert 2009, e.g., *Mit Ignatz Bubis starb ein deutscher Patriot*, lit. *with Ignatz Bubis died a German patriot*, 'The deceased Ignatz Bubis was a German patriot'), impersonal passives (e.g., *In meinem Zimmer wird geschlafen*, lit. *in my room is slept*, 'My room is for sleeping'), and dative experiencer constructions (e.g., *Mir friert die Nase*, lit. *to-me freezes the nose*, 'My nose is freezing'). In that connection, while we hope to have shown that the approach of exposing different populations of L2 learners to stimuli sentences in a common language

provides a useful tool for endeavors of this kind, it should be also pointed out that the present study involves genetically related and typologically similar languages, which are highly likely to display parallelisms in grammatical structure, to the extent that many of their constructions (especially the most basic ones) can be traced back to the proto-language these languages are descended from. Recruiting speakers of an L1 that is unrelated to, and/or typologically different from, the target language (here, German), could circumvent this issue, by allowing potential effects of constructional tolerance that would not be dwarfed by transfer effects of specific constructions (or at least less so).

Another aspect that should be explored in more detail is the role played by figurative language. Figurative uses are neither marginal nor clearly distinguishable from literal uses of language (Lakoff & Johnson 1980); in fact, they generally improve the communicative effectiveness of a message, which, notably, explains their omnipresence in political discourse (cf. Lakoff 1991; L'Hôte & Lemmens 2009). Along similar lines, Citron & Goldberg (in press) report neurological evidence that metaphorical uses of words are more emotionally engaging than their non-metaphorical counterparts. The particular status of figurativeness makes it an important source of creativity in language, which equally touches on the creative uses of constructions, and, consequently, on constructional tolerance. At the same time, as we mentioned regarding the reflexive-motion construction, some constructions seem more likely to be used figuratively than others, which bears on their degree of openness. In sum, it remains to be seen whether figurativeness is involved to the same extent in creative uses of all constructions and in different languages, and, in turn, how this parameter interacts with constructional tolerance.

References

Aijmer, K., & Hasselgård, H. (Eds.) (2004). *Translation and corpora*. Gothenburg: Acta Universitatis Gothoburgensis.

Alderson, J. C. (2006). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London/New York: Continuum.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390—412.

Bäckström, L., Lyngfelt, B., & Sköldberg, E. (2014). Towards interlingual constructicography: On correspondence between constructicon resources for English and Swedish. *Constructions and Frames*, 6(1), 9–33.

Bates, D., Maechler, M., & Ben Bolker (2011). lme4: Linear mixed-effects models using S4 classes. <http://CRAN.R-project.org/package=lme4>

Boas, H. C. (Ed.) (2010). *Contrastive studies in Construction Grammar*. Amsterdam/Philadelphia: John Benjamins.

Boas, H. C. (2011). Constructing parallel lexicon fragments based on English FrameNet entries: Semantic and syntactic issues. In H. Hedeland, T. Schmidt, & K. Woerner (Eds.), *Multilingual Resources and Multilingual Applications. Proceedings of the German Society for Computational Linguistics and Language Technology (GSCL) 2011, Hamburg* (pp. 9–18). University of Hamburg: Center for Language Corpora.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.

Citron, F., & Goldberg, A. E. (in press). Metaphorical expressions are more emotionally engaging than literal paraphrases. *Journal of Cognitive Neuroscience*.

Croft, W. (2001). *Radical Construction Grammar. Syntactic theory in typological perspective*. Oxford: Oxford University Press.

De Swart, H. E. (1998). Aspect Shift and Coercion. *Natural language and linguistic theory*, 16, 347-385.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of Let alone. *Language*, 64(3), 501-538.

Frense, J., & Bennett, P. (1996). Verb alternations and semantic classes in English and German. *Language Sciences*, 18(1-2), 305-317.

Fried, M., & Östman, J-O (Eds.) (2004). *Construction Grammar in a cross-language perspective*. Amsterdam: John Benjamins.

Gass, S. M., & Selinker, L. (1992). *Language transfer in language learning*. Amsterdam: John Benjamins

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.

Hilpert, M. (2009). The German *mit*-predicative construction. *Constructions and Frames*, 1(1), 29-55.

Hilpert, M. (2014). *Construction Grammar and its application to English*. Edinburgh: Edinburgh University Press.

Höder, S. (2012). Multilingual constructions: A diasystematic approach to common structures. In K. Braunmüller & C. Gabriel (Eds.), *Multilingual individuals and multilingual societies* (pp. 241–257). Amsterdam/Philadelphia: Benjamins.

Israel, M. (1996). The *Way* constructions grow. In A. Goldberg (Ed.), *Conceptual structure, discourse and language* (pp. 217-230). Stanford: CSLI.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. New York: Routledge.

Lakoff, G. (1991). Metaphor and war: The metaphor system used to justify war in the Gulf. *Journal of Urban and Cultural Studies*, 2(1), 59–72.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Theoretical prerequisites*. Stanford: Stanford University Press.

Lauwers, P. & Willems, D. (2011). Coercion: Definition and challenges, current approaches, and new trends. *Linguistics*, 49(6), 1219–1235.

L'Hôte, E., & Lemmens, M. (2009). Reframing treason: Metaphors of change and progress in new Labour discourse. *CogniTextes*, 3. URL : http://cognitextes.revues.org/248

Michaelis, L. A. (2005). Entity and event coercion in a symbolic theory of syntax. In J.-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive grounding and theoretical extensions* (pp. 45–87). Amsterdam/Philadelphia: John Benjamins.

Perek, F., & Goldberg, A. E. (submitted). Generalizing beyond the input: the functions of the constructions matter.

Proost, K. (2009). Warum man nach Schnäppchen jagen, aber nicht nach Klamotten bummeln kann: Die *nach*-Konstruktion zwischen Lexikon und Grammatik. *Opal* 4/2009, 10– 41. URL: http://pub.ids-mannheim.de/laufend/opal/pdf/opal09-4_proost.pdf

R Development Core Team (2012). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. URL: http://www.R-project.org/.

Rostila, J. (2014). Inventarisierung als Grammatikalisierung: produktive Präpositionalobjekte und andere grammatikalisierte Linking-Muster. In A. Lasch & A. Ziem (Eds.), *Grammatik als Netzwerk von Konstruktionen? Sprachliches Wissen im Fokus der Konstruktionsgrammatik* (pp. 127-153). Berlin/ New York: Walter de Gruyter.

Schütze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language*, 65, 1–14.

Wonnacott, E., Newport E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, 56, 165–209.

Ziem, A., & Lasch, A. (2013). *Konstruktionsgrammatik. Konzepte und Grundlagen gebrauchsbasierter Ansätze*. Berlin: De Gruyter.

Appendix 1: Stimuli sentences

Das Baby schreit nach der Flasche. (The baby screams after the bottle)

Das Huhn pickt nach den Körnern. (The chicken picks at the grains)

Das Kind kneift nach der Katze. (The child pinches at the cat)

Das Kind streckt nach der Schokolade. (The child stretches at the chocolate)

Das Lama spuckt nach den Besuchern. (The llama spits at the visitors)

Der Affe greift nach der Banane. (The monkey grips at the banana)

Der Angler schlägt nach einer Stechmücke. (The fisherman hits at the mosquito)

Der Detektiv fragt nach der Uhrzeit. (The detective asks after the time)

Der Fuchs schnüffelt nach der Wurst. (The fox sniffs after the sausage)

Der Gärtner gießt nach der Pflanze. (The gardner waters at the plant)

Der Gast verlangt nach der Rechnung. (The guest asks after the check)

Der Großvater fühlt nach der Türklinke. (The grandfather feels after the doorknob)

Der König schickte nach einem Arzt. (The king sends after a doctor)

Der Mann tastet nach dem Lichtschalter. (The man feels after the light switch)

Der Opa erkundigt nach der Enkelin. (The grandfather inquires after the granddaughter)

Der Papagei hackt nach der Erdnuss. (The parrot hacks at the peanut)

Der Polizist schießt nach dem Angreifer. (The cop shoots at the attacker)

Der Spieler tritt nach dem Ball. (The player kicks at the ball)

Die Abenteurer suchen nach dem Schatz. (The adventurers search after the treasure)

Die Angestellte denkt nach der Kollegin. (The employee thinks of her colleague)

Die Lehrerin bückt nach dem Stift. (The teacher bends after the pencil)
Die Leute rufen nach einem Taxi. (The people call after a taxi)
Die Mutter guckt nach dem Baby. (The mother watches after the baby)
Die Retter schaufeln nach den Verschütteten. (The rescuers shovel after the buried victims)
Der Torwart springt nach dem Ball. (The goalie jumps after the ball)

Der Dieb raubt sich ins Gefängnis. (The thief robs his way into prison)
Der Leser liest sich ins mittelalterliche Rom. (The reader reads his way into medieval Rome)
Der Mönch meditiert sich ins Nirvana. (The monk meditates his way into Nirvana)
Der Passagier steigt sich in den Zug. (The passenger enters his way onto the train)
Der Pensionär erinnert sich in seine Jugend. (The pensioner remembers his way into his youth)
Die Kinder trauen sich in das unheimliche Haus. (The children dare their way into the scary house)
Die Kleine flüchtet sich in die Arme der Mutter. (The small girl escapes her way into her mother's arms)
Die Neue integriert sich in die Gruppe. (The new member integrates her way into the group)
Die Sportler zwängen sich in die Gummianzüge. (The athletes force their way into the rubber suits)
Die Sportlerin turnt sich ins Finale. (The athlete exercises her way into the finals)
Die Straße führt sich in die Stadt. (The street leads its way into town)
Die Würmer beißen sich in die Darmwand. (The worms bite their way into the intestines)
Die Gruppe begibt sich in den Speisesaal. (The group makes its way into the dining hall)
Die Helfer wühlen sich durch den Schutt. (The helpers make their way through the rubble)
Die Pflanze wächst sich in das Fenster. (The plant grows its way into the window)
Die Ratte nagt sich in die Speisekammer. (The rat gnaws its way into the pantry)
Das Mädchen läuft sich ins Haus. (The girl runs her way into the house)
Der Astronaut schwebt sich ins Weltall. (The astronaut floats his way into space)
Der Bergsteiger arbeitet sich in die Höhe. (The mountaineer works his way to the top)
Der Trinker säuft sich ins Koma. (The alcoholic drinks his way into a coma)
Der Sportler boxt sich in die erste Liga. (The athlete boxes his way into the first league)
Der Tenor sang sich in die Hitparade. (The tenor sang his way into the charts)
Die Bagger graben sich in den Untergrund. (The diggers dig their way into the ground)
Der Schwimmer taucht sich ins Wasser. (The swimmer dives his way into the water)
Die Autorin schrieb sich in die Fantasie der Kinder. (The writer wrote her way into the children's minds)
Die Bakterien fressen sich in den Zahn. (The bacteria eat their way into the tooth)
Die D-Jugend schießt sich in die Tabellenspitze. (The youth team shoots its way into the top of the league)

Der Einbrecher droht das Opfer ins Badezimmer. (The burglar threatens the victim into the bathroom)
Der Gärtner füllt das Wasser in die Gießkanne. (The gardner fills the water into the watering can)
Der Hund bellt die Kinder ins Haus. (The dog barks the children into the house)
Der Müllmann greift den Sack in den Müllwagen. (The dumpster grips the bag into the dumpster truck)
Der Pfleger redet die Kranke auf die Liege. (The nurse talks the patient onto the bed)

Der Radfahrer fährt den Mann von dem Fußweg. (The bicyclist rides the man from the pathway)
Der Räuber steckte die Beute in einen Sack. (The robber put the bounty into a bag)
Der Redakteur bestellt den Gast in sein Büro. (The editor asks the guest into his office)
Der Soldat schleudert die Granate gegen das Tor. (The soldier hurls a grenade against the portal)
Der Verkäufer legt die Tüte auf die Waage. (The salesman puts the bag onto the scale)
Der Soldat sinkt das Boot in die Tiefe. (The soldier sinks the boat into the abyss)
Der Spieler wirft den Ball ins Tor. (The player throws the ball into the goal)
Die Arbeiter tragen die Bretter zur Baustelle. (The workers carry the planks to the site)
Die Ärztin bittet den Mann ins Wartezimmer. (The doctor asks the man into the waiting room)
Die Entführer zerrten das Opfer in einen Mercedes. (The kidnappers pulled the victim into a mercedes)
Die Fans schreien das Team ins Halbfinale. (The fans scream the team into the semi-finals)
Die Frau spaziert den Hund durch die Stadt. (The woman walks the dog through the town)
Die Retter brachten den Mann in eine Klinik. (The rescuers brought the man into a hospital)
Das Kind trinkt den Saft in den Bauch.(The child drinks the juice into his belly)
Das Kind zieht den Teddy hinter sich her. (The child pulls the teddy behind itself)
Das Mädchen schubst den Kater vom Sofa. (The girl pushes the cat from the couch)
Der Bauer pfeift die Hühner in den Stall. (The farmer whistles the chickens into the coop)
Der Chef verteilt das Geld an die Mitarbeiter. (The boss distributes the money to the employees)

Der Gärtner pflegt ihm den Garten. (The gardener maintains him the garden)
Der Klemptner schickt uns seinen Kollegen. (The plumber sends us his colleague)
Der Koch empfiehlt ihnen das Fischfilet. (The chef recommends him the fish fillet)
Der Moderator lacht ihm die Antwort. (The anchorman laughs him the answer)
Der Postbote bringt mir ein Päckchen. (The mailman brings me a package)
Der Professor denkt uns eine Erklärung. (The professor thinks us an explanation)
Der Richter verpasste ihm eine Geldstrafe. (The judge gave him a penalty)
Der Trainer nimmt ihm die Angst. (The coach takes him the fear)
Der Vater schenkt ihm eine Bohrmaschine. (The father gives him an electric drill)
Der Verkäufer sagt uns den Kaufpreis. (The salesman tells us the price)
Die Deutschen liefern uns das Werkzeug. (The Germans deliver us the tools)
Die Diebe stahlen ihr den Rucksack. (The thieves stole her the backpack)
Die Firma stellt ihm einen Dienstwagen. (The company gives him a company car)
Die Freundin telefoniert ihr die Neuigkeit. (The friend phones her the news)
Die Jugendlichen raubten ihm das Handy. (The youths robbed him the mobile phone)
Die Jury gab ihm zehn Punkte. (The jury gave him ten points)
Die Oma sponsort uns den Urlaub. (The grandmother sponsors us the vacation)
Die Stewardess wünscht uns einen guten Flug. (The stewardess wishes us a good flight)
Die Studentin erklärt ihm das Problem. (The student explains him the problem)
Das Ehepaar spendete ihnen tausend Euro. (The couple donated them thousand Euros)

Das Gericht erlaubte ihr die Abreise. (The court allowed her the departure)

Der Autor widmete ihr das Buch. (The author dedicated her the book)

Der Berater zeigte uns die Unterlagen. (The consultant showed us the papers)

Der Butler liest ihm die Zeitung. (The butler reads him the paper)

Der Chauffeur fährt ihnen die Tour. (The driver drives them the tour)

Appendix 2: DIALANG Vocabulary Size Placement Test (German version)

Dieser Test legt Ihnen eine Auswahl von 'Wörtern' vor, die zum Teil tatsächlich existieren, zum Teil aber frei erfunden sind. Bei diesen 'Wörtern' handelt sich um Verben, z.B. 'sprechen', 'laufen', 'essen', usw.

Wenn Sie der Meinung sind, dass das betreffende Wort tatsächlich existiert, so kreuzen Sie das Kästchen 'Ja' an. Wenn Sie der Meinung sind, dass es sich um ein erfundenes Wort handelt, so kreuzen Sie das Kästchen 'Nein' an.☐

schwören	☐ Ja	☐ Nein	hochjagen	☐ Ja	☐ Nein
zurückstecken	☐ Ja	☐ Nein	herausfinden	☐ Ja	☐ Nein
schuttern	☐ Ja	☐ Nein	krönen	☐ Ja	☐ Nein
umstellen	☐ Ja	☐ Nein	abstitzen	☐ Ja	☐ Nein
aggressieren	☐ Ja	☐ Nein	schlopfen	☐ Ja	☐ Nein
probieren	☐ Ja	☐ Nein	entfremden	☐ Ja	☐ Nein
stillen	☐ Ja	☐ Nein	lecken	☐ Ja	☐ Nein
fliehen	☐ Ja	☐ Nein	verfeinern	☐ Ja	☐ Nein
schmieren	☐ Ja	☐ Nein	wählen	☐ Ja	☐ Nein
abkratzen	☐ Ja	☐ Nein	herstellen	☐ Ja	☐ Nein
erdulden	☐ Ja	☐ Nein	einpfauen	☐ Ja	☐ Nein
schildieren	☐ Ja	☐ Nein	vertrenken	☐ Ja	☐ Nein
halbieren	☐ Ja	☐ Nein	leisten	☐ Ja	☐ Nein
bauen	☐ Ja	☐ Nein	geschehen	☐ Ja	☐ Nein
bremsen	☐ Ja	☐ Nein	orientieren	☐ Ja	☐ Nein
hineinbekommen	☐ Ja	☐ Nein	mögen	☐ Ja	☐ Nein
zermalmen	☐ Ja	☐ Nein	steuern	☐ Ja	☐ Nein
verhutzeln	☐ Ja	☐ Nein	hinstürzen	☐ Ja	☐ Nein
vereinlichen	☐ Ja	☐ Nein	umhaupten	☐ Ja	☐ Nein
seitern	☐ Ja	☐ Nein	umrahmen	☐ Ja	☐ Nein
festlegen	☐ Ja	☐ Nein	gemaunen	☐ Ja	☐ Nein
straübern	☐ Ja	☐ Nein	hinhauen	☐ Ja	☐ Nein
kleben	☐ Ja	☐ Nein	niederkämpfen	☐ Ja	☐ Nein
glasieren	☐ Ja	☐ Nein	leihen	☐ Ja	☐ Nein
pöhlen	☐ Ja	☐ Nein	beherrschen	☐ Ja	☐ Nein
bemollen	☐ Ja	☐ Nein	zählen	☐ Ja	☐ Nein
einarbeiten	☐ Ja	☐ Nein	entklupfern	☐ Ja	☐ Nein

zerdrücken	☐ Ja	☐ Nein	mindern	☐ Ja	☐ Nein
eindräuen	☐ Ja	☐ Nein	ansprechen	☐ Ja	☐ Nein
dreien	☐ Ja	☐ Nein	mitlaufen	☐ Ja	☐ Nein
binden	☐ Ja	☐ Nein	heraufsetzen	☐ Ja	☐ Nein
verderben	☐ Ja	☐ Nein	hineinstopfen	☐ Ja	☐ Nein
erinnern	☐ Ja	☐ Nein	plaschen	☐ Ja	☐ Nein
schweiken	☐ Ja	☐ Nein	entlaben	☐ Ja	☐ Nein
ausblintern	☐ Ja	☐ Nein	vergessen	☐ Ja	☐ Nein
chiffrieren	☐ Ja	☐ Nein	fordern	☐ Ja	☐ Nein
herauspasten	☐ Ja	☐ Nein	verballen	☐ Ja	☐ Nein
erkrellen	☐ Ja	☐ Nein			

---

[i] Here is the full list of items: a pencil, a saucepan, a bicycle, scissors, a cellphone, a rubbish bin, a cup, a hammer, a light bulb, a chair, a table, a vacuum cleaner, a bed, an envelope, a ruler, a newspaper, a beer crate, and an apple.

[ii] The practice items consisted of the following sentences: *Der Schnellzug fährt nach München* 'The express train goes to Munich', *Das Wetter regnet uns eine Katastrophe* 'The weather rains us a catastrophe', *Das Mädchen weint nach der Puppe* 'The girl cries at/for the doll'.

[iii] http://www.lancs.ac.uk/researchenterprise/dialang/about.htm, consulted February 8th 2013.

[iv] All tables presented in this paper use the following abbreviations: MCMCmean: Markov Chain Monte Carlo mean for the estimated coefficients; Lower, Upper: 95% highest posterior density intervals; pMCMC: *p*-value estimated by the sampling; Pr(>|t|): *p*-value based on the *t* distribution with the number of observations minus the number of fixed-effects coefficients as degrees of freedom.